⑲ 日本国特許庁（JP）

⑫ 公開特許公報（A）

⑪ 特許出願公開

昭59--111699

⑤Int. Cl.³
G 10 L 1/00

識別記号

庁内整理番号
R 7350--5D

㊸公開 昭和59年(1984)6月27日

発明の数 1
審査請求 未請求

（全 5 頁）

㊹話者認識方式

㉑特 願 昭57--221652
㉒出 願 昭57(1982)12月17日

㉒発 明 者 奈良泰弘
川崎市中原区上小田中1015番地
富士通株式会社内

㉒発 明 者 小林敦仁
川崎市中原区上小田中1015番地
富士通株式会社内

㉑出 願 人 富士通株式会社
川崎市中原区上小田中1015番地

㉔代 理 人 弁理士 山谷晧栄

明 細 書

1. 発明の名称

話者認識方式

2. 特許請求の範囲

人間が発音した音声を処理し，発音者の声が登録者の誰のものに似ているかを判定する話者認識装置において，複数の人間が発声した音声をフレーム周期で分析したパラメータ時系列を話者毎に保持する細分類音種パターンメモリを設け，発声者の音声をフレーム周期で分析したパラメータ時系列と細分類音種パターンメモリのパラメータ時系列との相関を演算してフレーム周期毎に最も相関の大きい登録話者名を選択する選択手段を設け，発声後最も多数回選択された登録話者名を決定してこれを話者認識結果として出力するようにしたことを特徴とする話者認識方式。

3. 発明の詳細な説明

発明の技術分野

本発明は話者認識方式に係り，特に複数の人間が発音した音声が登録されているときに入力音声がこの登録されている複数の人間の音声のうちどれともっとも類似しているものか認識できるようにしたものに関する。

技術の背景

現在の音声認識装置では，話者が自分の声で登録した辞書を使用した場合には高い認識率が得られるが，誰の声でも認識できるものではなく，他人の声で登録された辞書を使用した場合にはかなり認識率が低下する。したがつて，例えば電話で伝送された声により音声認識を行う場合には，第1段階として電話での話者が誰であるのか，あるいは複数登録話者のうちの誰に類似しているのかを認識する，話者認識が必要であつた。

従来技術と問題点

従来の話者認識法には「話し方」に着目する方法と，「音質」に着目する方法がある。前者は，例えば発声速度やイントネーションの変化パターンに注目する方法であるが，簡単な手法ではある

ものの，登録話者には音質に関連するデータが多く登録されており「話し方」のデータには音質に関する分析が不充分のため，音声認識装置の使用に先立つ辞書選びには適さない。また後者は発声者の声帯の形状や口腔等共鳴器官の形状により決まる音質に注目する方法であるが，既に登録してある複数話者のうちの１人が発声し，それが誰であるかを判定する用途には精度の点で不向きである反面，登録していない新らしい話者の声が誰のものに似かよっているかを判定するには好適である。

音質に注目した従来の話者認識技術には，発声音の音質をフレーム間隔毎に分析して特徴パラメータを抽出してからこの特徴パラメータを時間軸方向にたし合わせ平均化したものを話者毎に比較するというものがある。しかしこの方法では平均化したパターンに発声法の影響，すなわち調音の長短のような，音質の特徴以外に話し方の特徴がかなり含まれており，正確な認識には不充分であった。

3

発明の目的

本発明の目的はこのような問題を改善するために登録話者の音声をフレーム間隔毎に分析して得られるパラメータをメモリに格納しておき，入力音声の１フレーム毎に登録話者の誰の声に似ているかを判断し，入力音声の発声後に誰の声に似ているフレームが多かったかによって総合判断を行うことにより発声法の影響を受けずに，高精度に発声者の声が誰のものに類似しているかを判定できるようにした話者認識方式を提供することにある。

発明の構成

この目的を遂行するため，本発明の話者認識方式では，人間が発音した音声を処理し，発音者の声が登録者の誰のものに似ているかを判定する話者認識装置において，複数の人間が発声した音声をフレーム周期で分析したパラメータ時系列を話者毎に保持する細分類音韻パターンメモリを設け，発声者の音声をフレーム周期で分析したパラメータ時系列と細分類音韻パターンメモリのパラメータ時系列との相関を演算してフレーム周期毎に最も相関の大きい登録話者名を選択する選択手段を設け，発声後最も多数回選択された登録話者名を決定してこれを話者認識結果として出力するようにしたことを特徴とする。

4

発明の要点

本発明ではあらかじめ複数名の話者が発声した音声をフレーム間隔毎に分析して得られるパターン群を話者毎に整理してメモリ（細分類音韻メモリと呼ぶ）に格納しておく。そして話者認識すべき発声者の１フレームに対する分析結果と，細分類音韻メモリの全項目と相関（類似度）を計算し，最も類似度の高いパターンの発声者名を記録する。このような処理を話者認識すべき発声者の音声の全フレームに対して行ない，最も高い頻度で選択された発声者名を話者認識結果とするものである。

発明の実施例

本発明の一実施例を添付図面にもとづき詳述する。

図中，１はマイクロフォン，２は１６チャネルのバンドパス・フィルタ・バンク（以下バンドパス・フィルタという），３はマルチプレクサ，４はアナログ・ディジタル変換器（以下Ａ／Ｄ変換器という），５は細分類音韻メモリ，６はチェビシェフノルム計算回路，７は最小値演算部，８はデコーダ，９は登録話者類度記録部，１０は最大値演算部，S_1，S_2はそれぞれスイッチ部である。

バンドパス・フィルタ２はマイクロフォン１から入力された音声信号をf_1～f_{16}の１６の周波数に分析するものであって，スペクトルの概形を表わす１６チャネルのアナログ信号を出力するものである。

マルチプレクサ３は例えば１０msのサンプル周期毎に１回，チャネル１～１６のアナログ信号をスキャンすることにより時分割サンプルを行うものである。そしてこの時分割された１アナログ信号出力はＡ／Ｄ変換器４によりディジタル量に変換されて，例えば１６ワード／フレームのディジタル出力される。したがって入力発声長を例えば１秒間とすると，１発声について１００フレー

5

6

ム×16ワード=1600ワードが出力されることになる。

細分類音種メモリ5は登録者の特徴を保持するメモリでもつて，各登録者毎にその特徴を保持するために登録者毎にこれを用意する。したがつてこの例のように登録者が10名いる場合には細分類音種第1メモリ5－0～細分類音種第10メモリ5－9を用意する。

チエビシエフノルム計算回路6はフレームの類似度を計算するものでもつて

$$\sum_{i=1}^{16} |I_i - D_i|$$

を計算するものである。ここでI_iはA／D変換器4から出力される第iチヤネルを表わし，D_iはスイツチ部S_1を経由して伝達される細分類音種メモリ5に保持されている1辞書項目の第iワードを示す。この計算結果はA／D変換器4から送出される認識音声の1フレームデータ（16ワード）と，スイツチ部S_2を経由して細分類音種メモリ5から送出される1辞書項目（16ワード）の距離を表わすことになる。チエビシエフノルム計算回

路6は10m秒に1回，A／D変換器4から1フレーム分のデータが伝達されると，スイツチ部S_2を細分類音種第1メモリ5－0～細分類音種第10メモリ5－9個に順次切換え，100項目×10(組)=1000項目に対する距離計算を行うが，最小値演算部7はこの1000回の計算結果の最小値を演算し，その最小値を与えるデータが細分類音種第1メモリ5－0～細分類音種第10メモリ5－9のいずれから出力されたものであるかを示す4ビットの識別コードをフレーム毎に出力する。すなわち最小値演算部7は10m秒毎に1回，4ビットの識別コードを出力することになる。

デコーダ8はこの4ビットの識別コードを解読して，それが例えば細分類音種第1メモリ5－0から出力されたデータと比較したときに最小値が付与されたものであることを判読したとき，登録話者類度記録部9の第1カウンタ9－0に出力を送り，これを+1し，例えば細分類音種第2メモリ5－1から出力されたものと判読したとき第2カウンタ9－1に出力を送る。このようにして第

1カウンタ9－0～第10カウンタ9－9には各フレーム毎にもつとも類似した登録話者がカウントされることになり，これらのカウンタのうち最大値のものを最大値演算部10で検出することにより認識音声が，登録話者のどれともつとも類似しているのかを判別できる。

次に添付図面により本発明の動作を説明する。

(1)　登録時

登録時にはまずスイツチ部S_2を細分類音種第1メモリ5－0と接続し，第1番目の登録話者に例えばあらかじめ定められた音声を発音させる。この音声はマイクロフオン1から入力されてバンドパス・フイルタ2により16チヤネルに周波数分析され，16チヤネルのアナログ信号が出力される。マルチプレクサ3により10m秒のサンプル周期に1回チヤネル1～16のアナログ信号をスキヤンすることにより時分割サンプルを行ない，この出力はA／D変換器4によりデイジタル量に変換される。このようにしてA／D変換器4は10m秒毎に1チヤネル毎に1ワードの，合計し

て16ワードのデイジタル出力を生ずることになり，これが細分類音種第1メモリ5－0に登録されることになる。したがつて入力発声長が1秒の場合には，1発声について100フレーム×16ワード=1600ワードが登録されることになる。次に第2番目の登録話者が登録する場合，スイツチ部S_2を細分類音種第2メモリ5－1個に接続して同様の入力処理が行われるので，細分類音種第2メモリ5－1には第2番目の登録話者の特徴が保持される。このようなことが各登録話者毎に行われるので，登録話者が10名いるときには細分類音種第10メモリ5－9までに各登録話者の特徴が保持されることになる。

(2)　認識時

入力音声が登録話者の誰ともつとも類似しているかということを認識する場合には，スイツチ部S_2を開放状態にする。このとき入力される音声は，登録話者が細分類音種メモリ5に特徴を登録するときに発声したものと同じ音声であることが望ましい。マイクロフオン1から入力されたこの検認

識音声は，上記(1)と同様に16チャネルに周波数分析され，これらが10m秒のサンプル周期にスキャンされてデイジタル量に変換され，1フレーム16ワードのデイジタル出力がチエビシエフノルム計算回路6に伝達される。このときスイッチ部S_2は細分類音種第1メモリ5－0と接続して1ワードづつこのメモリの読出しを行ない，チエビシエフノルム計算回路6にて上記$\sum_{i=1}^{16}|I_i-D_{1i}|$で表現される計算を行う。すなわち被認識音声及び細分類音種メモリから得られた1項目16ワードのデータのそれぞれ対応する項の差の絶対値の和が計算されることになり，この計算結果がA／D変換器4から送られる1フレーム・データ（16ワード）と細分類音種メモリ5から送出される1辞書項目（16ワード）の距離を表わすことになる。チエビシエフノルム計算回路6は10m秒に1回，A／D変換器4から1フレーム分のデータが伝送されると，スイッチ部S_1を細分類音種第1メモリ5－0～細分類音種第10メモリ5－9側に順次切換えて，100項×10組に対する距離計

算を行うが，最小値演算部7はこの1000回の計算結果の最小値を演算してその最小値を与えるデータが細分類音種第1メモリ5－0～細分類音種第10メモリ5－9のいずれから出力されたものかを示す例えば4ビットの識別コードを出力する。すなわち最小値演算部7は10m秒に1回この識別コードを出力するが，この識別コードはデコーダ8で解読され，これに対応する第1カウンタ9－0～第10カウンタ9－9が選択的に+1される信号がデコーダ8より出力される。このようにして被認識音声の一発声が終つたとき，最大値演算部10はこの登録話者頻度記録部9を構成している第1カウンタ9－0～第10カウンタ9－9の値を比較して，その値も大きな値を示しているカウンタの番号を話者認識結果として出力するとともに，第1カウンタ9－0～第10カウンタ9－9をリセットする。

なお上記説明ではバンドパス・フイルタを16チャネルのものを使用した例について説明したが勿論このチャネル数はこれに限定されるものでは

なく適当なnチャネルにしたり，デイジタル・フイルタ・バンクを使用することもでき，またフレーム周期を10m秒ではなく他の適当な時間に変更することもできる。勿論登録話者は10人に限定されるものではなく任意の複数名に選定できる。また話者認識のときに発声する音声は，特定のものでも，登録時と認識時とが異なるものであつてもよい。

発明の効果

本発明によれば例えば語尾発声が長い短いというような発声法に影響されることなく，音質にもとづき話者認識を行うことができるので，高精度の話者認識を行うことができる。したがつて，これによりもつとも類似した登録話者の辞書を利用して不特定話者の音声認識率を高めることが可能となる。また電話を使用して入力される話者に対しても，本発明により前処理を行つて類似登録話者を選定し，その後にその登録辞書を使用することにより高精度の音声認識を行うことができる。

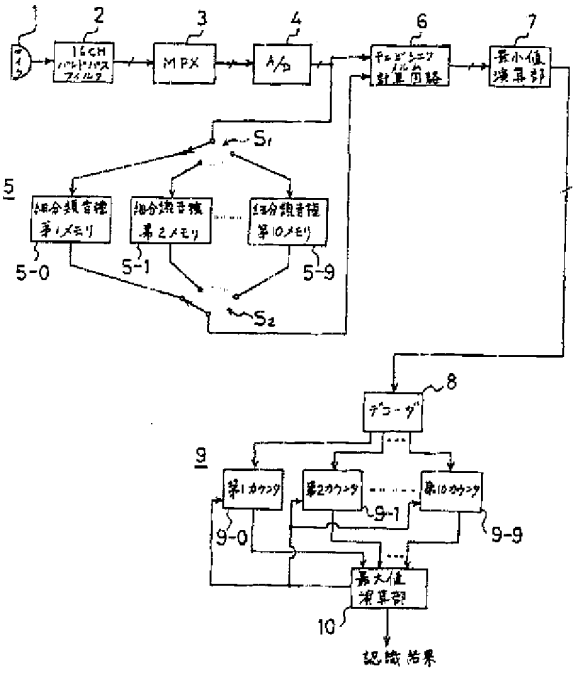
4　図面の簡単な説明

添付図面は本発明の一実施例構成図である。

図中，1はマイクロフオン，2はバンドパス・フイルタ・バンク，3はマルチプレクサ，4はアナログ・デイジタル変換器，5は細分類音種メモリ，6はチエビシエフノルム計算回路，7は最小値演算部，8はデコーダ，9は登録話者頻度記録部，10は最大値演算部，S_1，S_2はそれぞれスイッチ部である。

特許出願人　富士通株式会社

代理人 弁理士　山谷晴榮


1
2
3
4
6
7
MPX
A/D
S1
5
5-0
5-1
5-9
S2
8
デコーダ
9
9-0
9-1
9-9
10
認識結果